

# Sistema de compresión de voz mediante el estándar de compresión de imágenes JPEG2000

**Harold A. Romo Romero**

MSc. En Ingeniería, área Electrónica y Telecomunicaciones,  
Universidad del Cauca.

Docente Tiempo completo, Investigador Grupo GNTT  
Universidad del Cauca, Popayán, Colombia.  
[hromo@unicauca.edu.co](mailto:hromo@unicauca.edu.co)

**Mauricio Castillo Roza**

Ing. En Electrónica y Telecomunicaciones,  
Universidad del Cauca, Popayán, Colombia.

Investigador Grupo GNTT  
Universidad del Cauca, Popayán, Colombia.

**Gerardo Meneses**

Ing. En Electrónica y Telecomunicaciones,  
Universidad del Cauca, Popayán, Colombia.

Investigador Grupo GNTT  
Universidad del Cauca, Popayán, Colombia.  
[gmeneses@unicauca.edu.co](mailto:gmeneses@unicauca.edu.co)

**Resumen—** La amplia exploración de la teoría de wavelets ha dado lugar a importantes aplicaciones en el tratamiento de señales. Particularmente el análisis de señales de naturaleza aleatoria y no estacionaria, como la voz, es más adecuado si se realiza con transformaciones wavelets. Gracias a esta teoría, la definición del estándar de compresión de imágenes JPEG2000 se fundamenta en la transformada wavelet. Se propone la implementación de un sistema de compresión de voz mediante la técnica de compresión de imágenes definida en el estándar JPEG2000.

**Palabras clave—** Compresión de voz, JPEG2000, Procesamiento de imágenes, Transformada wavelet.

**Abstract—** The wide exploration of the wavelets theory has given place to important applications in the treatment of signs. Particularly the analysis of random and not stationary signs, as the voice, is more appropriate if it is carried out with wavelets transformations. Thanks to this theory, the definition of the standard of images compression JPEG2000 is based in the wavelet transform. Keeping in mind this, it's intends the implementation of a voice compression system by means of the images compression technique defined in the standard JPEG2000.

**Keywords—** Voice compression, JPEG2000, Image processing, Wavelet transform.

## I. INTRODUCCIÓN

La transformada wavelet como teoría matemática es relativamente reciente. Sin embargo, sus aplicaciones son ampliamente exploradas cada día,

desde la fotografía digital hasta aplicaciones en la industria editorial, geográfica o médica, entre otros. Particularmente en el ámbito tecnológico de las comunicaciones se destaca el análisis de señales e imágenes para efectos de compresión. Puntualmente, el Joint Pictures Expert Group (JPEG) desarrolló el estándar de compresión de imágenes basado en la transformada wavelet conocido como JPEG2000, el cual aventaja en rendimiento y capacidad de compresión a su formato predecesor, JPEG [2].

El codificador del estándar JPEG2000 se basa en una técnica de compresión con un banco de filtros que simulan una transformada wavelet discreta (Uniformly Maximally Decimated Filter Bank - UMDFB) [3]. Maneja una compresión con o sin pérdidas basada en una transformada wavelet traspuesta a una o varias matrices de referencia, fundamentándose en los esquemas de códigos de bloque embebidos de truncado optimizado (Embedded Block Coding with Optimized Truncation - EBCOT) [4-5].

En julio del 2005 en la Universidad de Santiago de Chile [5], Kaschel, trabajó en compresión de voz mediante la aplicación de formatos de compresión de imágenes basados en la transformada de coseno discreto como codificadores de voz, logró muy buenos resultados. Basados en ese

trabajo, aquí se propone un formato de compresión de voz, mediante la aplicación del formato de compresión de imágenes según la transformada *wavelet* JPEG2000.

A continuación se presenta un análisis del funcionamiento interno del estándar JPEG2000, se exponen los prototipos del sistema de compresión con sus respectivos objetivos de cumplimiento y las pruebas hechas al sistema con respecto a su rendimiento y capacidad, comparándolo con los esquemas tradicionales de codificadores de voz. Finalmente se presentan algunas conclusiones.

## II. ESTÁNDAR JPEG2000

JPEG2000 es un sistema de codificación de imágenes que usa técnicas de compresión basadas en la transformada *wavelet*. El proceso de compresión se implementa con o sin pérdidas de imágenes de uno o más componentes cromáticos. Adicionalmente, este estándar soporta diversas capacidades, tales como: 1) La recuperación progresiva de la imagen por fidelidad, resolución, posición o componentes; 2) la asignación de regiones de interés arbitrarias, donde la imagen se pueda codificar a diferentes niveles de resolución; 3) el acceso a cualquier segmento de la imagen sin tener que decodificar el archivo entero; 4) un formato de archivos flexible capaz de especificar información de opacidad o de secuencialidad en una imagen y 5) gran resistencia a los errores [1]. A continuación se presentan las especificaciones de la parte I del estándar, las que definen al sistema base [7].

JPEG2000 se basa en una variante de la transformada *wavelet* discreta (DWT), con cuantización escalar de límite nulo, modelación basada en contexto, codificación aritmética y asignación de tasa dinámica. La DWT es diádica y puede ser hecha por un filtro reversible, diseñado para la compresión sin pérdidas, o un filtro no reversible, que provee una mejor compresión pero con pérdidas de compresión. El cuantizador tiene una medida escalar con zona muerta y es independiente en cada sub-banda. Cada banda se divide en bloques de código, típicamente de 64x64, y su redundancia es minimizada con modelos de contexto y codificación aritmética de planos de bits. La información codificada se divide en capas, las cuales son niveles progresivos de la calidad de la imagen,

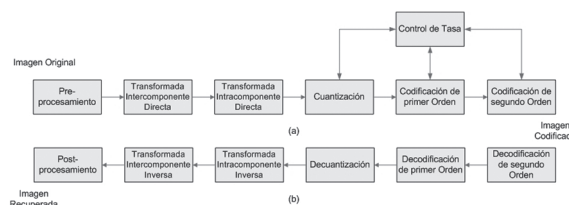
mediante la priorización en la asignación de tasa de bits dinámica.

Para explicar más coherentemente el sistema de compresión se muestra la estructura del codificador y decodificador de imágenes según el estándar del JPEG2000. El codificador pasa por varios procesos en cascada, cada uno tomando el subproducto entregado por el bloque anterior.

La Fig. 1 muestra el diagrama del codificador y del decodificador del estándar JPEG2000. Los bloques del codificador y decodificador son: *Pre-procesamiento/post-procesamiento*, *Transformada intercomponente*, *Transformada intracomponente*, *Cuantización/decuantización*, *Codificación de primer orden*, *Decodificación de segundo orden* y *Control de la tasa de compresión*.

Cada bloque del decodificador ha sido diseñado para invertir el proceso generado por el codificador, por lo que al mostrar el proceso de codificación se puede deducir el proceso inverso. Se supone que en el proceso ya se han efectuado los arreglos de formato necesarios.

FIG. 1 DIAGRAMA DE BLOQUES DEL ESTÁNDAR JPEG2000



Fuente: Los autores

## III. DISEÑO DEL SISTEMA DE COMPRESIÓN DE VOZ

Para el diseño del sistema de compresión de voz basada en el estándar JPEG2000, se investigaron los sistemas de referencia generados por el grupo JPEG, el cual, al realizar las recomendaciones del estándar JPEG2000, desarrolló dos sistemas software de referencia: *JasPer* [8] y *JJ2000* [9]. Estos sistemas software se explican en la parte V del estándar y fueron diseñados como modelos de verificación del estándar JPEG2000 [10].

Ambos sistemas se pueden utilizar como plataformas de desarrollo de la herramienta, ya que son la implementación explícita del estándar JPEG2000. Comparados los dos sistemas como posibles plataformas de desarrollo, se opta por el sistema *JJ2000* por su arquitectura orientada

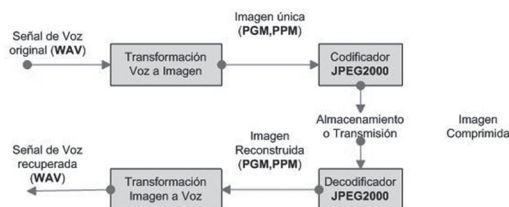
a objetos y su diferenciación funcional respecto a los módulos del estándar, como base para el diseño de sistema de compresión de voz JPEG2K\_VC aquí propuesto.

#### A. Diagrama en bloques del sistema JPEG2K\_VC

De acuerdo al funcionamiento del estándar JPEG2000, un diagrama en bloques del sistema de compresión de voz mediante técnicas de procesamiento digital de imágenes se presenta en la Fig. 2. Estos son descritos a continuación:

1. Transformación voz a imagen. Este bloque sirve de interfaz, convierte los datos del archivo de voz a una imagen equivalente para que el compresor de imágenes la pueda procesar. El formato wav es uno de los más utilizados para almacenar sonidos, guarda las muestras una tras otra sin ningún tipo de compresión de datos, con cuantificación uniforme. Su sencillez lo hace ideal para su procesamiento y manipulación. Este bloque debe analizar el archivo de voz para extraer del encabezado sus características y reorganizar las muestras como una imagen única equivalente.
2. Codificador JPEG2000. Este bloque procesa la imagen obtenida de la transformación del archivo de voz original y entrega un archivo de imagen comprimido, usa para ello las especificaciones propias del estándar JPEG2000.
3. Decodificador JPEG2000. Contrario al bloque anterior, este bloque se encarga de revertir los efectos de la compresión, recupera una imagen única a partir de los datos comprimidos.
4. Transformación de imagen a voz. Este último bloque retoma los datos del archivo de voz y reconvierte la imagen recuperada en el archivo de voz inicial.

FIG. 2 SISTEMA DE COMPRESIÓN DE VOZ JPEG2K\_VC



Fuente: Los autores

Durante el proceso de transformación de un archivo de voz a imagen y viceversa, se debe tener

en cuenta las características que definen a cada tipo de dato. Por un lado, las tramas de voz consisten en una serie de señales unidimensionales distribuidas en el tiempo a una tasa específica, cuyas muestras pueden tener uno o más bits de profundidad. Las imágenes, a su vez, son matrices de muestras de un bit de profundidad distribuidas espacialmente a una resolución específica. De acuerdo a las similitudes y diferencias entre ambos formatos se decidió por un proceso de interfaz simple, se separa la señal en sus componentes individuales y segmentándola posteriormente para ser insertada en la imagen equivalente.

#### B. Directrices de diseño del sistema JPEG2K\_VC

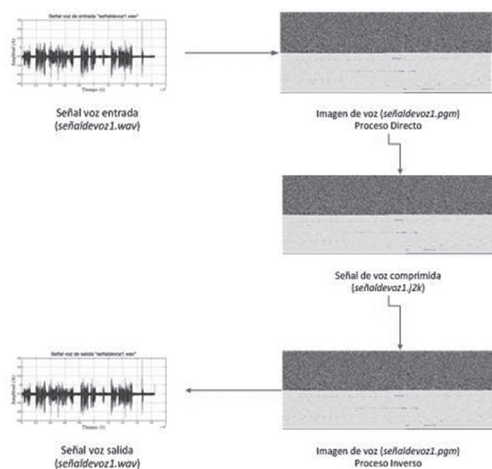
Del análisis de las tramas de voz e imagen y del trabajo con la herramienta de compresión, las directrices que guiaron el desarrollo del prototipo del sistema de compresión fueron:

1. Implementación de la interfaz. Este aspecto hace referencia al funcionamiento de la herramienta de compresión. Se plantearon dos opciones: una, modificar internamente la estructura del estándar de compresión JJ2000, de manera que pudiera manejar tramas de voz y dos, desarrollar un módulo de adaptación externo, de tal manera que el codificador funcione tal como está especificado en el estándar JJ2000. Por simplicidad y flexibilidad del sistema se prefiere la opción uno.
2. Organización de los datos a comprimir. Luego de trabajar con la herramienta se observa que la tasa de compresión mejora según las características propias de la imagen como el contraste y la correlación entre píxeles vecinos. Por esta razón se evita el simple mapeo de las muestras del archivo de voz en el arreglo matricial de la imagen y se procede a dividir internamente los datos de cada muestra por número de canales y profundidad de bit para lograr una imagen con mayor correlación entre bits vecinos y menor contraste global de la imagen, tal como se aprecia en la Fig. 3.
3. Procesamiento de encabezados. La información del encabezado en las tramas de voz identifica las principales características del archivo wav para su decodificación e inter-

pretación en recepción, por lo que no admite pérdidas de información. Para esto se identifica y utiliza un campo de información dentro del formato del estándar de compresión de imágenes JJ2000 que garantiza la conservación de la información aquí alojada. De esta manera se logra que se pueda recuperar siempre el archivo de voz en formato “wav” con todas sus características originales.

En la Fig. 3 se muestran los subproductos de cada módulo. Estos subproductos han sido obtenidos al elegir una compresión con pérdidas sin tasa asignada, con lo que el codificador asume una compresión mínima y una fidelidad máxima.

FIG. 3 SUBPRODUCTOS DEL PROTOTIPO JPEG2K\_VC



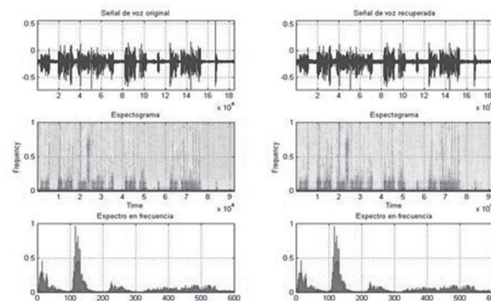
Fuente: Los autores

#### IV. RESULTADOS

Para validar la funcionalidad y eficiencia del sistema de compresión de voz mediante la aplicación del estándar JPEG2000, los archivos wav fueron sometidos al proceso de compresión para comparar: su tamaño con respecto a los archivos originales y su fidelidad después de su descompresión. En cada caso se varió la tasa de compresión, para determinar la tasa óptima de compresión que garantice una recuperación aceptable de la voz.

En la Fig. 4 se muestra la energía recuperada en la imagen, el análisis espectral de las señales de entrada y salida, muestra las frecuencias recuperadas de la señal y el espectrograma de cada una de ellas.

FIG. 4 ANÁLISIS ESPECTRAL DE LAS SEÑALES ENTRADA - SALIDA.



Fuente: Los autores

En la Tabla I se presenta para varias tasas de compresión (en bits por muestra bpm), el tamaño del archivo comprimido y el error cuadrático medio del archivo recuperado con respecto al archivo original. De esta forma se cuantifica la fidelidad del archivo de voz recuperado del sistema de compresión.

TABLA I  
ERROR CUADRÁTICO MEDIO VARIADA LA TASA DE COMPRESIÓN

Tasa de compresión	Tamaño del archivo comprimido (J2K)	Error cuadrático medio
8bpm	198KB	0.0311
7bpm	198KB	0.0311
6bpm	198KB	0.0311
5bpm	169KB	0.0316
4bpm	135KB	0.0321
3bpm	101KB	0.0324
<b>2bpm</b>	<b>68KB</b>	<b>0.0349</b>
1bpm	35KB	0.0450
0.5bpm	17KB	0.0477

Fuente: Los autores

En la Tabla II se presentan los resultados del análisis subjetivo del estándar con la puntuación de opinión media (Mean Opinion Score – MOS). Para ello, se pide la opinión de 20 encuestados sobre la calidad de voz de los archivos recuperados mientras se variaba la tasa de compresión, evaluando la calidad del archivo recuperado de 1 a 5.

TABLA II  
CALIDAD DE LA VOZ VARIADA LA TASA DE COMPRESIÓN

Tasa de compresión	Tamaño del archivo comprimido (J2K)	MOS
8bpm	686KB	5
7bpm	686KB	5
6bpm	662KB	5
5bpm	552KB	4.5
4bpm	451KB	4
3bpm	344KB	3.5
<b>2bpm</b>	<b>221KB</b>	<b>3.5</b>
1bpm	113KB	2.5
0.5bpm	55KB	1.5
0.1	11KB	1
0.05	6KB	1
0.01	1KB	1

Fuente: Los autores

Finalmente en la Tabla III se presentan los resultados de comparar el desempeño del sistema de compresión de voz aquí propuesto JPEG2K\_VC con otros codificadores de voz, ampliamente utilizados en sistemas de comunicaciones de voz.

TABLA III  
PUNTUACIONES MEDIA DE OPINIÓN PARA DIFERENTES CÓDEC

CÓDEC	Velocidad de transmisión de Datos [Kb/S]	MOS
GSM FR	12.5	3.5
G.723.1r53	5.3	3.65
G.729 <sup>a</sup>	8	3.7
G.726 ADPCM	32	3.8
GSM EFR	12.2	3.8
G.723.1r63	6.3	3.9
G.723.1r63	8	3.92
iLBC	15.2	4.14
AMR	12.2	4.14
G.711 (RDSI)	64	4.3
JPEG2K_VC	Variable	3.5

Fuente: Los autores

## V. CONCLUSIONES

Las valoraciones objetiva y subjetiva del proceso de compresión y descompresión de archivos de voz, mediante la aplicación del estándar de compresión de imágenes JPEG2000 evidencian la viabilidad y eficiencia del sistema aquí propuesto, pues su desempeño es comparable con el de los CÓDEC existentes que permite una compresión hasta del 70% del tamaño original de un archivo de voz.

El proceso de adaptación y conversión de archivos de voz a imagen es relativamente sencillo, lo que posibilita su aplicación directa en las emergentes redes de VoIP. Su característica adicional de permitir tasas de compresión adaptables, lo hace ideal para permitir adecuar la transmisión de VoIP a las condiciones cambiantes del canal de comunicaciones.

En el futuro se puede pensar en desarrollar un estándar de compresión de voz basado en wavelets, sin recurrir al estándar JPEG2000 como mediador, de esta manera se evita la transformación de archivos wav a formatos de imagen reconocibles por el JPEG2000.

## REFERENCIAS

- [1] Adams Michael D. "The JPEG-2000 Still Image Compression Standard". Dept. of Electrical and Computer Engineering University of Victoria. 3 Diciembre 2005
- [2] "ISO/IEC call for contributions, JPEG 2000," ISO/IEC JTC 1/SC 29/WG 1 N505, Marzo 1997.
- [3] W. Sweldens, "The lifting scheme: A construction of second generation wavelets". SIAM Journal of Mathematical Analysis, vol. 29, no. 2, pp. 511-546, Mar. 1998
- [4] B.E. Usevitch. "A Tutorial on Modern Lossy Wavelet Image Compression: Foundations of JPEG2000". IEEE Signal Processing Magazine. Vol. 18. pp. 22-35. Septiembre 2001
- [5] D. Taubman, "High performance scalable image compression with EBCOT". IEEE Trans. on Image Processing, vol. 9, no. 7, pp. 1158-1170, Julio 2000
- [6] Kaschel Héctor C. Francisco Watkins, Enrique San Juan U. "Compresión de voz mediante técnicas para el procesamiento de señales y aplicación de formatos de compresión de imágenes". Universidad de Santiago de Chile, Departamento de Tecnologías Industriales. 14 de octubre de 2005.

- [7] ISO/IEC 15444-1: Information technology-JPEG 2000 image coding system-Part 1: Core coding system, 2000
- [8] Adams Michael D. and F. Kossentini, "JasPer: A software-based JPEG-2000 codec implementation". IEEE International Conference on Image Processing, Vancouver, BC, Canada, Octubre 2000.
- [9] Raphaël Grosbois, Pierre Gerbelot, Touradj Ebrahimi, "Authentication and access control in the JPEG 2000 compressed domain". SPIE 46th Annual Meeting, Applications of Digital Image Processing XXIV, Vol. 4472, pp. 95-104, San Diego, July 29th-August 3rd, 2001
- [10] ISO/IEC 15444-5: Information technology-JPEG 2000 image coding system-Part 5: Reference software, 2002